Systems and Methods for Big and Unstructured Data Project

Author(s): **Gabriele Ginestroni**

Giacomo Gumiero

Lorenzo Iovine

Nicola Landini

Francesco Leone

Group Number: **10**

Academic Year: 2022-2023

Contents

1 | Introduction

In this chapter will be presented the problem specification and the hypothesis under which the database is implemented.

1.1. Problem Specification

This project aims to build an Information System that handles scientific articles contained in the DBLP bibliography. The project involves managing the type of every article and the associated DOI (Digital Object Identifier), which identifies an article or a document and links to it on the web. Other entities to deal with are authors, identified by an ID or ORCID (Open Researcher and Contributor ID), and their affiliations with organizations. In order to address the problem, we will store data in a graph database, allowing us to visualize relations and handle information correctly.

1.2. Assumptions

- All the data in the dataset are heterogeneous, so fields are different
- The **authors** with missing field _id are not considered
- It is possible that an author writes for different organizations
- Field _id in **author** exists and it is unique
- Field _id in **article** exists and it is unique
- It is impossible that 2 different articles are on the same journal, in the same *volume* with an intersection between *page_start* and *page_end*
- The designed model doesn't take into consideration the URL associated to the article node, as the main focus of the project was not reading the article
- It is possible to find a self-reference in a publication
- A *venue* can be instantiated as a journal, a conference or a generic venue

1.3.   Figures, Tables and Algorithms

Figures, Tables and Algorithms have to contain a Caption that describe their content, and have to be properly reffered in the text.

1.3.1.   Figures

For including pictures in your text you can use `TikZ` for high-quality hand-made figures, or just include them as usual with the command

`\includegraphics[options]{filename.xxx}`

Here xxx is the correct format, e.g. `.png`, `.jpg`, `.eps`, ...

Figure 1.1: Caption of the Figure to appear in the List of Figures.

Thanks to the `\subfloat` command, a single figure, such as Figure 1.1, can contain multiple sub-figures with their own caption and label, e.g. Figure 1.2a and Figure 1.2b.

(a) One PoliMi logo.   (b) Another one PoliMi logo.

Figure 1.2: This is a very long caption you don't want to appear in the List of Figures.

1.3.2.   Tables

Within the environments `table` and `tabular` you can create very fancy tables as the one shown in Table 1.1.

Title of Table (optional)

	column 1	column 2	column 3
row 1	1	2	3
row 2	α	β	γ
row 3	alpha	beta	gamma

Table 1.1: Caption of the Table to appear in the List of Tables.

You can also consider to highlight selected columns or rows in order to make tables more readable. Moreover, with the use of `table*` and the option `bp` it is possible to align them at the bottom of the page. One example is presented in Table 1.2.

	column1	column2	column3	column4	column5	column6
row1	1	2	3	4	5	6
row2	a	b	c	d	e	f
row3	α	β	γ	δ	ϕ	ω
row4	alpha	beta	gamma	delta	phi	omega

Table 1.2: Highlighting the columns

	column1	column2	column3	column4	column5	column6
row1	1	2	3	4	5	6
row2	a	b	c	d	e	f
row3	α	β	γ	δ	ϕ	ω
row4	alpha	beta	gamma	delta	phi	omega

Table 1.3: Highlighting the rows

1.3.3. Algorithms

Pseudo-algorithms can be written in LaTeX with the `algorithm` and `algorithmic` packages. An example is shown in Algorithm 1.1.

Algorithm 1.1 Name of the Algorithm

 1: Initial instructions
 2: **for** $for-condition$ **do**
 3:      Some instructions
 4:      **if** $if-condition$ **then**
 5:          Some other instructions
 6:      **end if**
 7: **end for**
 8: **while** $while-condition$ **do**
 9:      Some further instructions
10: **end while**
11: Final instructions

1.4. Theorems, propositions and lists

1.4.1. Theorems

Theorems have to be formatted as:

Theorem 1.1. *Write here your theorem.*

Proof. If useful you can report here the proof.

1.4.2. Propositions

Propositions have to be formatted as:

Proposition 1.1. *Write here your proposition.*

1.4.3. Lists

How to insert itemized lists:

- first item;

- second item.

How to insert numbered lists:

1. first item;

2. second item.

A | Appendix A

If you need to include an appendix to support the research in your thesis, you can place it at the end of the manuscript. An appendix contains supplementary material (figures, tables, data, codes, mathematical proofs, surveys, ...) which supplement the main results contained in the previous chapters.

List of Figures

List of Tables